POLYPUBLIE

One sequence to segment them all: efficient data augmentation for CT and MRI cross-domain 3D spine segmentation

Nathan Molinier, Hendrik Möller, Thomas Dagonneau, Anna Curto, Robert Graf, Matan Atad, Daniel Rueckert, Jan S. Kirschke, Julien Cohen-Adad

Document en libre accès dans PolyPublie
Open Access document in PolyPublie

Date	2026
Type	Communication de conférence / Conference or Workshop Item Communication écrite/Paper
Version	Version finale avant publication / Accepted version
URL de PolyPublie **PolyPublie URL**	https://publications.polymtl.ca/83082/
Évaluation par les pairs **Peer reviewed**	Révisé par les pairs / Refereed
Conditions d'utilisation **Terms of use**	© Tous droits réservés / All rights reserved
Référence **Citation**	Molinier, N., Möller, H., Dagonneau, T., Curto, A., Graf, R., Atad, M., Rueckert, D., Kirschke, J. S., & Cohen-Adad, J. (septembre 2026). One sequence to segment them all: efficient data augmentation for CT and MRI cross-domain 3D spine segmentation [Communication écrite]. 29th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2026), Strasbourg, France. https://doi.org/10.1007/978-3-032-38072-2_38

Publication officielle
Official publication

Nom de la conférence **Conference name**	29th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2026)
Date et lieu **Date and location**	2026-09-27 - 2026-10-01, Strasbourg, France
Maison d'édition **Publisher**	Springer Nature
URL officielle **Official URL**	https://doi.org/10.1007/978-3-032-38072-2_38
Mention légale **Legal notice**	

Ce fichier a été téléchargé à partir de PolyPublie, le dépôt institutionnel de Polytechnique Montréal
This file has been downloaded from PolyPublie, the institutional repository of Polytechnique Montréal
https://publications.polymtl.ca

This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

One Sequence to Segment Them All: Efficient Data Augmentation for CT and MRI Cross-Domain 3D Spine Segmentation

Nathan Molinier[1,2,*], Hendrik Möller[3,4,*], Thomas Dagonneau[1,2], Anna Curto-Vilalta[4,5], Robert Graf[3,4], Matan Atad[3,4], Daniel Rueckert[3,4], Jan S. Kirschke[3], and Julien Cohen-Adad[1,2]

[1] NeuroPoly Lab, Polytechnique Montréal, Montréal, Québec, Canada
[2] Mila - Québec Artificial Intelligence Institute, Montréal, Québec, Canada
[3] Dept. for Interventional and Diagnostic Neuroradiology, TUM University Hospital
[4] Chair for AI in Healthcare and Medicine, Technical University of Munich (TUM)
[5] Dept. of Orthopedics and Sports Orthopedics, TUM University Hospital
[*]These authors contributed equally to this work.
{julien.cohen-adad,nathan.molinier}@polymtl.ca,hendrik.moeller@tum.de

Abstract. Deep learning–based medical image segmentation is increasingly used to support clinical diagnosis and develop new treatment strategies. However, model performance remains limited by the scarcity of high-quality annotated data and insufficient generalization across imaging protocols. This limitation is particularly evident in MRI and CT, where models are typically trained on a single acquisition sequence and exhibit reduced robustness when applied to unseen sequences or contrasts. Although data augmentation is widely used to improve general robustness on medical images, its impact on cross-modality generalization has not been quantitatively explored. In this work, we study a targeted set of data augmentation techniques designed to improve cross-modality transfer. We train three spine segmentation models, each on a single-modality/sequence dataset, and evaluate them across seven out-of-distribution datasets (spanning CT and MRI), reflecting a realistic single-sequence training and multi-sequence/contrast/modality deployment scenario. Our results demonstrate substantial performance gains on unseen domains (average Dice gain of 155%) while preserving in-domain accuracy (average Dice decrease of 0.008%), including effective transfer between CT and MRI. To mitigate the computational cost typically associated with strong data augmentation, we implement GPU-optimized augmentations that maintain, and even improve, training efficiency by approximately ∼10%. We release our approach as an open-source toolbox (GitHub), enabling seamless integration into commonly used frameworks such as nnUNet and MONAI. These augmentations significantly enhance robustness to heterogeneous clinical imaging scenarios without compromising training speed.

Keywords: medical segmentation · data augmentation · domain generalization · spine.

1 Introduction

Medical image segmentation is essential for computer-aided diagnosis and treatment development, enabling quantitative analysis and disease monitoring [14]. However, manual segmentation is costly, time-consuming, and prone to variability, motivating the use of automatic deep learning–based methods [16].

Despite strong performance, most models are tailored to specific imaging modalities/contrasts and often fail when applied to data from different modalities (CT, MRI), scanners (vendor, field strengths, software version) or protocols. This issue is particularly pronounced in spine segmentation, where methods are typically modality-specific [23]. Although CT and MRI provide complementary information, the underlying anatomy remains consistent, with domain shifts mainly arising from contrast and intensity differences.

Data augmentation is widely used to improve generalization by increasing data variability. Numerous techniques have been proposed, from basic geometric and intensity transformations to synthetic data generation [1,8,10,15,20,24]. However, results are highly dataset- and task-dependent, and many studies conflate augmentation effects with architectural, dataset changes or inclusion of mixed sequences during training [4,12,19,2]. Furthermore, controlled analyses in 3D segmentation are scarce, with most work focused on 2D settings [6,18]. Finally, some methods proposed in the literature are difficult to adopt in practice, as they require auxiliary training, specialized pipelines, and can be computationally expensive [1,9].

In this work, we investigate a combination of data augmentations for 3D spine segmentation using nnUNet as a standardized baseline [11]. Models are trained on single-sequence datasets and evaluated on out-of-distribution data spanning multiple modalities (CT, MRI), contrasts (e.g. T1w, T2w, Dixon in-phase) and sequences (e.g., resolution, field-of-view). Our contributions are as follows:

- We propose a set of data augmentation techniques for domain generalization across CT and MRI sequences for spinal segmentation.
- While traditional frameworks rely on CPU for data augmentation, we present a software solution that performs data augmentation on GPU for nnUNet [11] and MONAI [3] frameworks.

2 Methodology

2.1 Dataset

We use three datasets comprising seven imaging sequences: the public Spider lumbar spine MRI dataset [7] and two private in-house datasets (SG and MM) spanning different modalities (CT, MRI), scanner manufacturers, imaging sequences, and spinal regions (cervical, thoracic, lumbar, sacral). Table 1 shows more details about the datasets, and Figure 1 illustrates representative examples. Informed consent was waived by the local ethics committee (27/19 S-SR). All datasets include manual reference segmentations from different expert raters.

Table 1. The demographics of the utilized datasets.

Property	MM	Spider [7]	SG
N subjects	240	218	144
Sequences	CT	MRI (T1w, T2w)	CT, MRI Dixon (in-phase, fat, water)
Spine Coverage	Whole spine	Lumbar only	Various parts
Sex (% female)	39	63	45
Age range (yrs)	31 – 87	14 – 84	20 – 94
Mean Age ± SD (yrs)	64 ± 11	60 ± 15	66 ± 19
Is public	✗	✓	✗
Date range	2005 – 2022	2019 – 2020	2017 – 2023
Pathology	Multiple myeloma	Lower back pain	Diverse
Manufacturers	Siemens, Philips	Siemens, Philips	Siemens, Philips, Canon, Toshiba, GE

Datasets were preprocessed as follows: all reference segmentation masks were re-labeled to contain only three semantic classes: vertebrae, intervertebral discs, and spinal canal. Images were re-oriented to (Posterior, Inferior, Right)-order and resampled to a 1 mm^3 isotropic resolution.

SG dataset was split into 80/10/10 subject-wise for training, validation, and testing, respectively. The Spider and MM datasets were used exclusively for evaluation. Training on Spider, which contains only lumbar scans, while the other datasets include a wider range of fields of view, introduces additional factors related to differences in spine coverage that would confound our analysis. The MM dataset is much more diverse, with subjects having multiple CT images from different scanners and noise levels, enabling more sophisticated out-of-distribution evaluation. Thus, we refrain from training on them and instead create a random 10/90 validation/test split for both datasets. The validation set for the latter datasets was used during development.

2.2 Data Augmentation

We use the nnUNet framework [11], a widely adopted framework for medical image segmentation. To address concerns that extensive data augmentation may slow training and hinder adoption, we developed a custom GPU nnU-Net. This trainer reproduces the default nnU-Net augmentation pipeline while extending

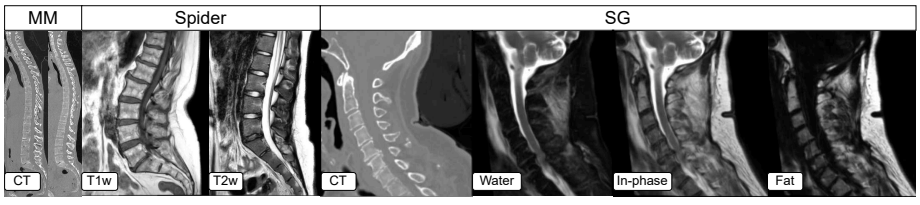


Fig. 1. Example images from the different datasets. In order left to right: MM (CT), Spider (MRI T1w, T2w) [7] and SG (CT, MRI Dixon water, in-phase, and fat).

it with additional GPU-based transformations that modify image appearance and simulate variations in imaging modality (CT vs. MRI) and contrast (e.g., T1- vs. T2-weighted MRI). The selected transformations are described in Table 2 and shown in the Figure 2.

Table 2. Summary of 3D data augmentation transforms used in this study.

Transform	Description
Intensity inversion [10]	Min-Max intensity range inversion
Scharr filtering [17]	Gradient magnitude enhancement for edge emphasis
RedistributeSeg [22]	Segmentation-driven regional intensity redistribution by adding a randomly scaled probability density function of voxel intensities within each segmented region.
RandomConv [24]	Random 3D convolutions to induce texture variation
Histogram equalization [22]	Estimate the intensity histogram, compute its normalized cumulative distribution function, and use it to remap voxel intensities.
Bias field [21]	Simulating intensity shading inhomogeneity
Unsharp masking [5]	High-frequency reinforcement for contour sharpening
Function transform	Apply a randomly selected nonlinear function

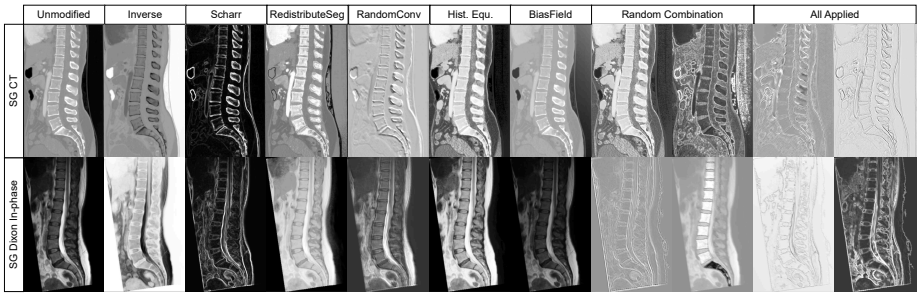


Fig. 2. Example transforms applied to one SG CT (top) and SG Dixon in-phase image (bottom) of the same subject. The columns are the different transforms applied to these images.

Standard geometric transformations (e.g., rotations and flips) and other baseline augmentations (e.g., Gaussian noise, blurring, and resampling) are kept with the same configuration as the normal nnUNet trainer. The order of the transforms is set to geometric transformations first, then the new augmentations described in Table 2, and finally the remaining default augmentations from the normal nnUNet trainer. We applied each augmentation with specific parameters and probabilities (for the exact configuration, see our configuration).

2.3 Experiments

Models were trained using a single sequence from the SG dataset. Validation only used in-sequence samples, whereas evaluation was performed on the test splits of all sequences/modalities described in the Dataset section. We set the batch size to 2 and the patch size to (128, 128, 128). We trained for 1,000 epochs with the default nnUNetPlans, corresponding to the PlainConvUNet architecture [11].

Our custom trainer was compared under the same data setting against the standard nnUNetTrainer baseline and the official nnUNetTrainerDA5 configuration, which represents the most extensive default nnUNet augmentation strategy. All models were evaluated quantitatively on the corresponding test sets for each sequence.

To quantify the contribution of individual augmentations, we performed ablation experiments in which each newly added augmentation was enabled with an application probability of 0.5, while disabling all other new augmentations. Additionally, disabling all of the baseline transformations highlights the difference our new augmentations make. Finally, we trained once while shuffling the order in which all but the geometric transformations are applied.

Segmentation performance was measured using the Dice similarity coefficient (per class). Metrics were computed with Panoptica [13]; the configuration we used for metric computation is available (see our configuration). We recorded the training time using nnUNet's internal timing and we averaged it across runs to compare the computational cost of our proposed GPU augmentation pipeline against the conventional nnUNet trainer.

We assessed the statistical significance using the Wilcoxon signed-rank test; $p < 0.05$ was considered significant.

For software and their versions, we use Python 3.11, nnUNetv2 v2.6.2, and panoptica v1.1.4. We compute the Wilcoxon signed-rank with SciPy v1.14.1. We managed the data using TPTBox v0.4.2. Our GPU implementation heavily leans on kornia v0.8.2.

3 Results

3.1 Segmentation performance

Table 3 shows how our custom trainer compares against the two nnUNet setups, the nnUNetTrainer (Base) and nnUNetTrainerDA5 (DA5). Our augmentation setup shows superior Dice performance on out-of-domain sequences across all three training setups. We observe this performance gain not only across the three semantic classes on average, but also quite consistently within each of them (see Table 4). Notably, the best overall boosts are for CT-MRI transfers, in both directions. On average across the three benchmarks and datasets, we reach an out-of-domain CT Dice performance of 0.83 when training on MRI. This is ∼91% of the respective average in-domain CT performance. Respectively, when training on CT, we reach an overall out-of-domain MRI performance that is ∼85% of the respective in-domain MRI Dice scores. For qualitative results,

Table 3. Three training setups (rows) evaluated on the test splits across all seven sequences (columns) of the datasets, comparing the common nnUNet trainers with ours. Values are the global dice score averaged across classes, then averaged across subjects. Bold indicates the best-performing value across trainers. An asterisk indicates statistical significance against the best other setup.

		MM	Spider		SG			
Trained on	Setup	CT	T1w	T2w	CT	fat	in-phase	water
SG in-phase	Base	0.13 ±0.09∗	0.76 ±0.04∗	**0.83 ±0.03**	0.17 ±0.13∗	0.79 ±0.06∗	**0.91 ±0.02**	0.82 ±0.08∗
SG in-phase	DA5	0.05 ±0.07∗	0.77 ±0.02∗	**0.83 ±0.02**	0.12 ±0.14∗	0.82 ±0.05∗	**0.91 ±0.02**	0.82 ±0.07∗
SG inphase	Ours	**0.83 ±0.03**∗	**0.82 ±0.02**∗	**0.83 ±0.02**	**0.83 ±0.05**∗	**0.87 ±0.03**∗	0.90 ±0.03∗	**0.86 ±0.03**∗
SG CT	Base	**0.91 ±0.01**	0.15 ±0.14∗	0.05 ±0.06∗	**0.91 ±0.03**	0.33 ±0.15∗	0.15 ±0.07∗	0.05 ±0.05∗
SG CT	DA5	**0.91 ±0.01**	0.26 ±0.17∗	0.12 ±0.10∗	**0.91 ±0.03**	0.51 ±0.09∗	0.21 ±0.10∗	0.13 ±0.08∗
SG CT	Ours	0.90 ±0.01∗	**0.79 ±0.03**∗	**0.78 ±0.03**∗	**0.91 ±0.03**	**0.75 ±0.07**∗	**0.77 ±0.05**∗	**0.71 ±0.07**∗
SG fat	Base	0.31 ±0.11∗	**0.81 ±0.02**	0.81 ±0.02	0.41 ±0.15∗	**0.89 ±0.03**∗	0.87 ±0.03	0.60 ±0.20∗
SG fat	DA5	0.31 ±0.12∗	**0.81 ±0.02**	0.81 ±0.02	0.45 ±0.17∗	**0.89 ±0.03**∗	0.87 ±0.03	0.56 ±0.20∗
SG fat	Ours	**0.83 ±0.02**∗	**0.81 ±0.02**	**0.82 ±0.02**	**0.83 ±0.05**∗	0.88 ±0.03∗	**0.88 ±0.03**	**0.84 ±0.04**∗

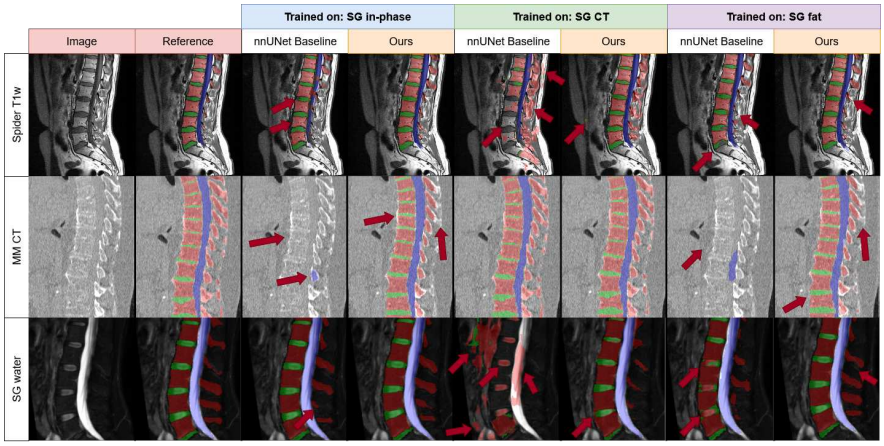


Fig. 3. A qualitative comparison between predictions produced by the baseline and our setup on random test samples. The rows represent different samples, labeled with the dataset name and sequence. The columns display the ground-truth segmentation of the test image, along with the predicted segmentations from both the baseline and our method, for each of the three training sets. Red arrows highlight areas with prediction errors.

Table 4. Class-wise global Dice score differences between three training setups comparing our method with the baseline nnUNetTrainer. Positive values indicate performance gains, while negative values indicate loss; asterisks denote statistical significance.

		MM	Spider		SG			
Trained on	Class	CT	T1w	T2w	CT	fat	in-phase	water
SG in-phase	Vertebra	+0.70∗	0.00∗	−0.01∗	+0.59∗	+0.02∗	−0.01∗	+0.04∗
SG in-phase	IVD	+0.70∗	+0.01∗	0.00	+0.60∗	0.00	−0.01∗	+0.09∗
SG in-phase	Canal	+0.70∗	+0.17∗	0.00	+0.79∗	+0.22∗	−0.01∗	−0.01∗
SG CT	Vertebra	−0.01∗	+0.62∗	+0.64∗	−0.01∗	+0.43∗	+0.50∗	+0.59∗
SG CT	IVD	−0.03∗	+0.64∗	+0.69∗	0.00∗	+0.52∗	+0.52∗	+0.57∗
SG CT	Canal	0.00	+0.64∗	+0.86∗	0.00	+0.32∗	+0.83∗	+0.82∗
SG fat	Vertebra	+0.69∗	−0.01∗	−0.01∗	+0.57∗	−0.01∗	0.00	+0.18∗
SG fat	IVD	+0.68∗	+0.02∗	+0.01	+0.59∗	−0.01∗	+0.01	+0.39∗
SG fat	Canal	+0.19∗	−0.01∗	+0.01∗	+0.10∗	−0.01∗	+0.01	+0.14∗

see Figure 3. For in-domain samples, our setup sometimes yields slightly lower Dice scores (avg. −0.008 across classes and training sets). Notably, baselines trained on SG Dixon in-phase or fat images perform well on Spider T1w and T2w images. We hypothesize that this is due to lumbar scans being easier to segment than other areas of the spine, as well as to T2w images being the same contrast as Dixon in-phase images.

The results of our ablations are in Table 5. They confirm that the proposed augmentations are the main contributors to the observed out-of-distribution gains, with all newly introduced transformations showing statistically significant effects ($p < 0.001$).

3.2 Training speed

All models were trained on an Nvidia A40 GPU. On average across the SG sequences, the default nnUNetTrainer baseline required ∼52 s per epoch (∼14.5 h total), while the nnUNetTrainerDA5 setup required ∼59 s per epoch (∼16.4 h total). Re-implementing the base setup with our GPU optimizations reduces this to ∼43 s per epoch (∼12 h total), which corresponds to a ∼17% speedup. When using all our proposed augmentations, training time becomes ∼47 s per epoch (∼13 h total), retaining a ∼10% speedup (saving ∼1.5 h). Inference time remains unchanged since the architecture is identical.

4 Discussion

Our results demonstrate that a carefully selected set of data augmentations significantly enhances cross-domain robustness at minimal cost. Unlike prior approaches [1,9], our method requires no auxiliary training or specialized pipelines. Thanks to GPU optimization, the more complex augmentation scheme does not increase training time and causes negligible, if any, in-domain performance loss.

Table 5. Ablation evaluation using the models trained on SG in-phase and SG CT images. The first column indicates the changes to our setup. The values are global dice scores averaged across classes, then across subjects. Bold values are the best single-applied transformations for each testing set.

	MM	Spider		SG			
Setup	CT	T1w	T2w	CT	fat	in-phase	water
Trained on: SG in-phase							
Ours	0.83	0.82	0.83	0.83	0.87	0.90	0.86
Ours (Base disabled)	0.77	0.81	0.83	0.81	0.87	0.90	0.86
Ours (Random order)	0.83	0.81	0.83	0.83	0.87	0.90	0.86
Base	0.13	0.76	0.83	0.17	0.79	0.91	0.82
Base + Intensity inversion	0.52	0.79	**0.83**	0.60	0.82	**0.91**	0.84
Base + Scharr filtering	0.65	0.81	**0.83**	0.62	0.84	**0.91**	0.85
Base + RedistributeSeg	**0.82**	**0.82**	**0.83**	**0.80**	**0.86**	**0.91**	**0.86**
Base + Random Conv	0.72	0.81	**0.83**	0.75	**0.86**	**0.91**	**0.86**
Base + Histogram equalization	0.45	0.78	**0.83**	0.43	0.81	**0.91**	0.83
Base + Bias field	0.44	0.77	**0.83**	0.41	0.80	**0.91**	0.83
Base + Unsharp masking	0.36	0.76	**0.83**	0.31	0.78	**0.91**	0.83
Base + Function transform	0.40	0.77	**0.83**	0.38	0.81	**0.91**	0.83
Trained on: SG CT							
Ours	0.90	0.79	0.78	0.91	0.75	0.77	0.71
Ours (Base disabled)	0.90	0.77	0.76	0.91	0.75	0.75	0.52
Ours (Random order)	0.90	0.79	0.78	0.90	0.75	0.77	0.72
Base	0.91	0.15	0.05	0.91	0.33	0.15	0.05
Base + Intensity inversion	**0.91**	0.57	0.48	**0.91**	0.49	0.31	0.09
Base + Scharr filtering	0.90	0.57	0.35	**0.91**	0.53	0.25	0.07
Base + RedistributeSeg	**0.91**	**0.75**	**0.73**	**0.91**	**0.67**	0.57	0.15
Base + Random Conv	0.90	0.72	0.68	**0.91**	0.65	**0.68**	**0.27**
Base + Histogram equalization	**0.91**	0.31	0.15	**0.91**	0.40	0.21	0.09
Base + Bias field	**0.91**	0.39	0.19	**0.91**	0.48	0.23	0.08
Base + Unsharp masking	**0.91**	0.38	0.19	**0.91**	0.43	0.21	0.07
Base + Function transform	**0.91**	0.36	0.18	**0.91**	0.39	0.20	0.09

Among these, *RedistributeSeg* yields the largest benefits for cross-modality and cross-contrast transfer by exploiting the segmentation mask to independently adjust intensities inside and outside the region of interest, thereby generating localized intensity shifts that better capture modality-specific variations. Texture-oriented transformations, such as *Scharr* and *Random Convolution*, further enhance generalization by increasing textural diversity and more closely approximating appearance changes induced by modality or contrast differences. While our ablation may suggest that individual augmentations can match the performance of combined ones, models trained on SG CT and evaluated on SG water images reveal exceptions, where single transformations achieve a maximum Dice score of 0.27 compared to 0.71 when all augmentations are combined. These results underscore that no single augmentation is universally optimal for MRI or CT, and suggest that the synergistic effect of combining multiple augmentations is essential for achieving the best performance. Although different

transformation orderings produce visually distinct samples, randomly shuffling the order does not lead to significant performance gains.

Several limitations should be acknowledged. Segmentations across datasets were generated by different annotators, and MRI–contrast–dependent tissue boundaries introduce ambiguity in the reference labels, contributing to variability. Moreover, spinal canal boundaries are not directly visible in CT, rendering their annotation inherently probabilistic. Nevertheless, these annotations remain sufficiently informative to highlight segmentation failures relative to the nnUNet baseline, as shown in Figure 3. Finally, our evaluation is limited to semantic spine segmentation, and the generalizability of these findings to other tasks remains to be established.

Future work will investigate optimized augmentation strategies that further leverage segmentation masks, extend the evaluation to additional segmentation tasks, and compare our method with prior approaches [1,9] to provide a more comprehensive assessment of its effectiveness and generalizability. To facilitate reproducibility and adoption, we release our GPU-based augmentation framework as open source (GitHub) with native support for nnUNet and MONAI. Such augmentations can improve model performance across heterogeneous clinical imaging settings while shifting manual annotation efforts toward rare pathologies that are difficult to synthesize, rather than simple intensity-based variations.

Disclosure of Interests. The authors have no competing interests in the paper

References

1. Billot, B., Greve, D.N., Puonti, O., Thielscher, A., Van Leemput, K., Fischl, B., Dalca, A.V., Iglesias, J.E., et al.: Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining. Medical image analysis **86**, 102789 (2023)
2. Bédard, S., Karthik, E.N., Tsagkas, C., Pravatà, E., Granziera, C., Smith, A., Weber II, K.A., Cohen-Adad, J.: Towards contrast-agnostic soft segmentation of the spinal cord. Medical Image Analysis (2025)
3. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: Monai: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
4. Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., Haworth, A.: A review of medical image data augmentation techniques for deep learning applications. Journal of medical imaging and radiation oncology **65**(5), 545–563 (2021)
5. Deng, G.: A generalized unsharp masking algorithm. IEEE transactions on Image Processing **20**(5), 1249–1261 (2010)
6. Goceri, E.: Medical image data augmentation: techniques, comparisons and interpretations. Artificial intelligence review **56**(11), 12561–12605 (2023)
7. van der Graaf, J.W., van Hooff, M.L., Buckens, C.F., Rutten, M., van Susante, J.L., Kroeze, R.J., de Kleuver, M., van Ginneken, B., Lessmann, N.: Lumbar spine segmentation in mr images: a dataset and a public benchmark. Scientific Data **11**(1), 264 (2024)

8. Graf, R., Möller, H., McGinnis, J., Rühling, S., Weihrauch, M., Atad, M., Shit, S., Mühlau, M., Paetzold, J.C., Rueckert, D., et al.: Modeling the acquisition shift between axial and sagittal mri for diffusion superresolution to enable axial spine segmentation. In: Medical Imaging with Deep Learning (2024)
9. Graf, R., Schmitt, J., Schlaeger, S., Möller, H.K., Sideri-Lampretsa, V., Sekuboyina, A., Krieg, S.M., Wiestler, B., Menze, B., Rueckert, D., et al.: Denoising diffusion-based mri to ct image translation enables automated spinal segmentation. European Radiology Experimental **7**(1), 70 (2023)
10. Häntze, H., Xu, L., Rattunde, M.N., Donle, L., Dorfner, F.J., Hering, A., Nawabi, J., Adams, L.C., Bressem, K.K.: Mri annotation using an inversion-based preprocessing for ct model adaptation. European Radiology Experimental **9**(1), 93 (2025)
11. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
12. Kebaili, A., Lapuyade-Lahorgue, J., Ruan, S.: Deep learning approaches for data augmentation in medical imaging: a review. Journal of imaging **9**(4), 81 (2023)
13. Kofler, F., Möller, H., Buchner, J.A., de la Rosa, E., Ezhov, I., Rosier, M., Mekki, I., Shit, S., Negwer, M., Al-Maskari, R., et al.: Panoptica–instance-wise evaluation of 3d semantic and instance segmentation maps. arXiv preprint arXiv:2312.02608 (2023)
14. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
15. Ouyang, C., Chen, C., Li, S., Li, Z., Qin, C., Bai, W., Rueckert, D.: Causality-inspired single-source domain generalization for medical image segmentation. IEEE Transactions on Medical Imaging **42**(4), 1095–1106 (2022)
16. Rayed, M.E., Islam, S.S., Niha, S.I., Jim, J.R., Kabir, M.M., Mridha, M.: Deep learning for medical image segmentation: State-of-the-art advancements and challenges. Informatics in medicine unlocked **47**, 101504 (2024)
17. Saydazimov, J., Ergashev, S., Nosirkulov, A.: Research of some image filter algorithms used in object detection. In: Proceedings of the 8th International Conference on Future Networks & Distributed Systems. p. 781–785. Association for Computing Machinery, New York, NY, USA (2025). https://doi.org/10.1145/3726122.3726236, https://doi.org/10.1145/3726122.3726236
18. Schwonberg, M., El Bouazati, F., Schmidt, N.M., Gottschalk, H.: Augmentation-based domain generalization for semantic segmentation. In: 2023 IEEE Intelligent Vehicles Symposium (IV). pp. 1–8. IEEE (2023)
19. Schwonberg, M., Gottschalk, H.: Domain generalization for semantic segmentation: A survey. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6437–6448 (2025)
20. Shin, H.C., Tenenholtz, N.A., Rogers, J.K., Schwarz, C.G., Senjem, M.L., Gunter, J.L., Andriole, K.P., Michalski, M.: Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: International workshop on simulation and synthesis in medical imaging. pp. 1–11. Springer (2018)
21. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4itk: improved n3 bias correction. IEEE transactions on medical imaging **29**(6), 1310–1320 (2010)

22. Warszawer, Y., Molinier, N., Valosek, J., Benveniste, P.L., Bédard, S., Shirbint, E., Mohamed, F., Tsagkas, C., Kolind, S., Lynd, L., Oh, J., Prat, A., Tam, R., Traboulsee, A., Patten, S., Lee, L.E., Achiron, A., Cohen-Adad, J.: Totalspineseg: Robust spine segmentation with landmark-based labeling in mri. ResearchGate preprint (2025), https://www.researchgate.net/publication/389881289_TotalSpineSeg_Robust_Spine_Segmentation_with_Landmark-Based_Labeling_in_MRI
23. Xie, Z., Lin, Z., Sun, E., Ding, F., Qi, J., Zhao, S.: Deep learning for automatic vertebra analysis: A methodological survey of recent advances. Computerized Medical Imaging and Graphics p. 102652 (2025)
24. Xu, Z., Liu, D., Yang, J., Raffel, C., Niethammer, M.: Robust and generalizable visual representation learning via random convolutions. arXiv preprint arXiv:2007.13003 (2020)